HUMANITARIAN ACADEMY · HUMANITAIRES JOBS

Mini MBA AI and Humanitarian Action: working with Claude and Gemini

Course handout · Week 3 : Protecting data: what you never give an AI, and how to anonymise

A programme of exactly 3 months (12 weeks, 144 hours) to use the AI assistants Claude (Anthropic) and Gemini (Google) in humanitarian work, in a useful, safe and responsible way: how they work and their limits, writing requests, data protection, ethics and humanitarian principles, reports, proposals, data analysis and MEAL, translation and communication with communities, research and source checking, disinformation and security, productivity and automation, use policy and deployment in an organisation. Corrected exercises, three monthly exams with case studies, and a verifiable diploma from the Humanitarian Academy. Independent programme, with no partnership with Anthropic or Google.

https://humanitairejobs.org/en/academy/mini-mba-ia-claude-gemini/semaine-3/

Free document from the Humanitarian Academy. It may be printed and shared freely for training purposes, without modification.

Week 3 : Protecting data: what you never give an AI, and how to anonymise

Objectives

- Apply the principles of data protection in humanitarian action to the use of AI assistants.
- Classify information (public, internal, confidential, sensitive) and decide whether it can be given to an assistant, and with which plan.
- Anonymise or pseudonymise a document before using it with AI, and adjust privacy settings.

Why this is the most important week

Humanitarian organisations hold some of the most sensitive data there is: identities of refugees or displaced people, survivors of violence, separated children, health data, team locations, security incidents, information about armed groups. A leak can expose people to reprisals, discrimination or violence.

When you paste text into an AI assistant, that text **leaves your computer** and is processed on the provider's servers, often in another country. Depending on the plan and settings, it may be **retained**, sometimes **read by people** in charge of quality, or used to **improve the models**.

> Principle: never give an AI assistant data that you would not be allowed to email to an external contractor without a data processing agreement.

The reference principles

The **Handbook on Data Protection in Humanitarian Action** by the ICRC and the Brussels Privacy Hub (2nd edition, 2020, which also covers new technologies), the data protection policies of UN agencies and, for the organisations concerned, the European Union's General Data Protection Regulation (GDPR) set out common principles:

- **lawfulness and purpose**: data is collected for a specific purpose and is not reused for an incompatible purpose;
- **minimisation**: use only the data strictly necessary;
- **security**: protect data against unauthorised access;
- **transfers**: regulate any transfer to a third party or another country;
- **rights of individuals**: information, access, rectification.

Giving the data of a beneficiary list to an AI assistant for the general public is a **transfer to a third party**, for a **purpose** that was not announced to the people concerned. In the vast majority of cases, this is contrary to these principles.

The sector also speaks of **demographically identifiable data**: even without names, aggregated data (for example the precise location of an ethnic or religious minority group) can expose an entire community.

What you never give a consumer assistant

- Names, phone numbers, card or registration numbers, photos, addresses, GPS coordinates of beneficiaries or staff members.
- Any information about survivors of gender-based violence, children, protection cases, mental health or HIV, even roughly "anonymised".
- Detailed security incident reports, routes, positions and travel times, evacuation plans, information on access negotiations.
- Allegations of sexual exploitation and abuse, fraud or corruption under investigation.
- Individual health data.
- Passwords, access keys, bank details.
- Documents marked confidential by the organisation, a donor or a partner.

For these categories, even a professional plan with contractual commitments is not always enough: a documented decision by the organisation (data protection impact assessment) is required before any use.

Consumer plans and professional plans

In November 2026, the broad lines are as follows (always check the terms in force):

- **Consumer plans (free or individual subscription)**: both providers offer settings to control history retention and the use of conversations to improve models. At Google, Gemini Apps activity may be retained and conversations may be reviewed by people; a setting allows it to be turned off. At Anthropic, the user chooses in the privacy settings whether their conversations may be used for training, and Claude offers an incognito chat mode.
- **Plans for organisations** (Claude Team and Enterprise; Gemini in Google Workspace for business and education): providers commit contractually not to use customer data to train their models by default, with central administration, access controls and data processing terms.

In practice:

- a free personal account should only be used for **public or non-sensitive** information;
- **internal** (non-sensitive) information belongs in a professional plan approved by the organisation;
- **personal and sensitive** data remains excluded, unless there is a formal, regulated decision.

> Turning off history or training use reduces risks, but does not turn a personal account into an authorised tool for beneficiary data.

Classify, then anonymise

Before each use, classify the information according to the organisation's policy, for example:

- **Public**: published reports, calls for proposals, legal texts. Usable with any assistant.
- **Internal**: draft proposals, procedures, minutes without personal data. Approved professional plan.
- **Confidential**: detailed budgets, partner assessments, negotiations. Professional plan and manager's approval, after removing unnecessary elements.
- **Sensitive**: personal data, protection, security, investigations. Excluded.

Anonymising means making it impossible to identify a person, including by cross-referencing. **Pseudonymising** means replacing identifiers with codes ("Beneficiary A"); the person remains identifiable by whoever holds the lookup table: this is still personal data.

Common traps: a village of 300 inhabitants plus "34-year-old woman, widowed, four children" is enough to identify someone; a focus group quote may contain a name or nickname; file metadata (author, comments, revisions) contain names. Remove unnecessary details, group ages and places, and delete comments and document history.

Good everyday habits

- Use the **account provided by the organisation** for work, never a personal account for internal documents.
- Check the **privacy settings** at first login, then every quarter: history, training use, memory, enabled connectors.
- Limit **connectors** (access to email, Drive) to what is strictly necessary: an assistant that can read your whole mailbox can also read sensitive data in it.
- Do not share **public links** to conversations containing internal information.
- In the field, watch out for **visible screens** and shared devices; log out.
- In case of error (sensitive data pasted by mistake): delete the conversation, then **report the incident** to the data protection focal point without delay. An error reported quickly limits the damage; a hidden error makes it worse.

Exercise: can AI be used for this document?

Situation. A protection team wants to save time on five tasks: A) summarise the public report of a protection cluster; B) translate into English the minutes of an internal meeting without names; C) analyse a spreadsheet of 2,300 households with names, phone numbers and vulnerabilities to find duplicates; D) rephrase the account of a survivor of violence for a case report; E) propose a staff training plan on PSEA. The organisation has a professional plan approved by the data protection officer, with a commitment not to use data for training.

Task. For each task: allowed with a personal account, allowed with the professional plan only, or excluded? Justify.

Answer. A) Allowed with any account: public document. B) Professional plan only: internal non-sensitive document, after checking that it really contains no names or security elements. C) Excluded in this form: personal and vulnerability data of 2,300 households; deduplication is done in the organisation's beneficiary management tool, or the AI can be asked to write a deduplication formula or script tested on fictitious data, then run locally. D) Excluded: gender-based violence data, extremely sensitive, even with a professional plan. E) Allowed with any account if the request contains no sensitive internal information; preferably with the professional plan in order to attach internal procedures.

Key takeaways

- Text pasted into an assistant leaves the organisation: it is processed, sometimes retained, by the provider.
- Minimisation, purpose and regulated transfers apply to the use of AI.
- Never put personal data of beneficiaries, protection cases, health or security data into a consumer assistant.
- Professional plans offer contractual commitments, but do not replace a decision by the organisation for sensitive data.
- Pseudonymising is not anonymising; an error is reported immediately.

Quiz

1. Pasting a beneficiary list into a consumer assistant is:
   A. A transfer to a third party for an unannounced purpose, to be avoided
   B. Allowed if history is turned off
   C. Allowed if the list is in English
   D. Without consequence

2. Replacing names with "Beneficiary A, B, C" while keeping a lookup table is:
   A. Anonymising
   B. Encrypting
   C. Deleting
   D. Pseudonymising: this is still personal data

3. A detailed security incident report with routes and times:
   A. Can be summarised by a free account
   B. Must not be given to an assistant without a formal decision by the organisation
   C. Can be given if it is translated
   D. Must be published

4. You pasted sensitive data by mistake. What should you do?
   A. Nothing
   B. Continue the conversation
   C. Delete the conversation and report the incident to the focal point without delay
   D. Change browser

5. The principle of minimisation means:
   A. Writing short requests
   B. Using as few tokens as possible
   C. Paying for the cheapest plan
   D. Using only the data strictly necessary

Quiz answers : 1. A · 2. D · 3. B · 4. C · 5. D